

# Properties of Kinetic Transition Networks for Atomic Clusters and Glassy Solids

John W. R. Morgan,<sup>a</sup> Dhagash Mehta,<sup>b</sup> and David J. Wales<sup>\*c</sup>

Received Xth XXXXXXXXXX 20XX, Accepted Xth XXXXXXXXXX 20XX

First published on the web Xth XXXXXXXXXX 20XX

DOI: 10.1039/b000000x

A database of minima and transition states corresponds to a network where the minima represent nodes and the transition states correspond to edges between the pairs of minima they connect via steepest-descent paths. Here we construct networks for small clusters bound by the Morse potential for a selection of physically relevant parameters, in two and three dimensions. The properties of these unweighted and undirected networks are analysed to examine two features: whether the networks are small-world, where the shortest path between nodes involves only a small number of edges; and whether the networks are scale-free, having a degree distribution that follows a power law. Small-world character is present, but statistical tests show that a power law is not a good fit, so the networks are not scale-free. These results for clusters are compared with the corresponding properties for the molecular and atomic structural glass formers ortho-terphenyl and binary Lennard-Jones. These glassy systems do not show small-world properties, suggesting that such behaviour is linked to the structure-seeking nature of the Morse clusters.

## 1 Introduction

The potential energy surface (PES)<sup>1</sup> of an atomic cluster corresponds to the energy as a function of the coordinates specifying the configuration. The most interesting points on the surface are usually the local minima and transition states, which are stationary points where the gradient of the potential is zero. For local minima, the potential energy rises for any infinitesimal displacement of internal coordinates, while for transition states there is a unique negative Hessian (second derivative matrix) eigenvalue.<sup>2</sup> Treating the PES as a network can provide insight into the overall structure of the energy landscape. Here the network in question is formed by considering minima as the nodes and transition states as edges between the minima they connect via steepest-descent paths.<sup>3,4</sup> Two key questions are then whether the network is small-world and scale-free.

The degree of a node is the number of edges connected to it. The degree distribution is then defined as the number of nodes with given degrees.<sup>5</sup> A path is a sequence of nodes connected by edges, with the length of the path being the number of edges it contains. The shortest path length gives the minimum number of edges between a pair of nodes, and the average shortest path length is the shortest path length averaged over all pairs of nodes. The clustering coefficient is the ratio of the number of connections between neighbours of a node to

the number of such connections that could exist.<sup>6</sup>

Two useful references are provided by models based on a lattice graph and a random graph. Random graphs have a small clustering coefficient and a slowly growing average shortest path length.<sup>5,7</sup> In contrast, lattice graphs, which are composed of a regular array of nodes with edges only between nearest-neighbours in space, have a comparatively large average shortest path length and a larger clustering coefficient.<sup>6</sup> Small-world networks were introduced by Watts and Strogatz<sup>6</sup> and defined as networks that show a high degree of local clustering behaviour, similar to a lattice graph, but also a short path length, even between distant nodes, as exhibited by random graphs. The name comes from experiments performed by Milgram with letter passing, which demonstrated that most citizens of the USA are separated from each other by a surprisingly small number of social contacts.<sup>8</sup> This result is now known in terms of “six degrees of separation” from the estimated average path length.<sup>9</sup> Watts and Strogatz showed that networks from a wide variety of areas show small-world behaviour, including the neural network of *C. elegans*, the power grid of the Western USA, and the collaborations of film actors (linked to the concept of the Bacon number).<sup>10</sup>

A scale-free network has a degree distribution with a power law tail, as defined by Barabási and Albert.<sup>11,12</sup> This distribution implies there are a small number of nodes with a very high degree, called hubs, and more nodes with a smaller degree, acting as local hubs, down to nodes with only a few connections, in a hierarchical fashion. Power laws were fitted for a variety of available networks, including the World Wide Web and the film actor collaboration graph, and a preferential at-

<sup>a</sup> Department of Chemical Engineering, University of Michigan, Ann Arbor, MI 48109-2136, USA. E-mail: [jwrm@umich.edu](mailto:jwrm@umich.edu)

<sup>b</sup> Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, IN 46556, USA. E-mail: [dmehta@nd.edu](mailto:dmehta@nd.edu)

<sup>c</sup> University Chemical Laboratories, Lensfield Road, Cambridge CB2 1EW, UK. E-mail: [djw34@cam.ac.uk](mailto:djw34@cam.ac.uk)

tachment model was suggested to explain how the behaviour arises. Following this work, power laws were fitted to many other networks, such as the interactions between solar magnetic loops leading to a solar flare,<sup>13</sup> and aftershocks of earthquakes.<sup>14</sup> However, few of these studies used robust statistical methods to determine whether or not a power law was a good fit to the data, and many of the results have been called into question by Clauset *et al.*<sup>15</sup>

Doye and Massen<sup>16,17</sup> studied Lennard-Jones (LJ) clusters and concluded that the networks were both small-world and scale-free. Small-world behaviour is believed to have implications for self-assembly, as it suggests that the global minimum can be reached from anywhere on the PES by a relatively short transition pathway. One would therefore expect that systems with a funnel-like PES, characteristic of structure-seekers, may display small-world behaviour.<sup>18</sup> Scale-free behaviour was more surprising, as there is no obvious analogy to the addition of new nodes in the preferential attachment model, the most straightforward route to a power law distribution. However, noticing a correlation between minima with a high degree and a low potential energy, it was suggested that due to the larger basins of attraction for low energy minima, any other minimum was more likely to be connected to a low energy minimum than a high energy one, thus establishing an analogy to preferential attachment. This possibility was further investigated using Apollonian packings,<sup>19</sup> concluding that the contacts between discs in the 2D Apollonian packing form a scale-free network with a spatial distribution that may resemble the catchment basins of a PES.

Other energy landscapes have been studied with a view to analysing the same properties. Protein folding networks have been considered by Rao and Caffisch.<sup>20</sup> They used molecular dynamics to generate snapshots of the structure and then derived a conformation based on the secondary structure each residue belonged to. These snapshots do not precisely correspond to minima on the PES, so the formation of the network is not equivalent to the LJ results considered above, but the idea is similar. A power law was fitted to the degree distribution with the conclusion was that the network was scale-free, drawing attention to the hubs present in the network.

Bowman and Pande studied a Markov state model for another protein folding landscape.<sup>21</sup> They noted that native states were hubs and suggested that this was characteristic of a scale-free network, but without consideration of the degree distribution. Similarly, Chakraborty *et al.* considered an RNA folding network, noting that it may be scale-free due to the presence of hubs, but without providing the degree distribution.<sup>22</sup> It is, however, possible to construct networks with hubs that are not scale-free, the simplest example being the star network,<sup>5</sup> in which there is one central node to which all others are attached and there no other edges. Therefore, although scale-free networks do contain hubs, the presence of

hubs is not in itself enough evidence to conclude that a network is scale-free. Recently, Mehta *et al.*<sup>23</sup> constructed stationary point databases for the Thomson problem<sup>24</sup> up to 150 particles. They demonstrated that the networks show small-world behaviour, and that the presence of hubs suggested the possibility of the networks being scale-free, but acknowledged that further statistical testing was required. Recently, an analysis of networks of minima for machine learning problems has also been initiated.<sup>25</sup>

Potential energy surfaces for glassy systems have a large number of competing low-lying amorphous minima with similar energies, separated by high barriers.<sup>1</sup> Here we consider two examples: bulk ortho-terphenyl (OTP), which is often represented by a coarse-grained model consisting of a LJ site at the centre of each benzene ring;<sup>26</sup> and bulk binary Lennard-Jones (BLJ), in which there are two different types of particles, interacting via LJ potentials with different parameters.<sup>27,28</sup> If the small-world behaviour observed for atomic clusters is the result of a funnelled landscape, then we anticipate that glassy systems will be different.

The Morse potential<sup>29</sup> is a widely used pairwise representation for modelling atomic clusters.<sup>30–37</sup> The functional form is

$$V_M(R) = \epsilon e^{\rho(1-R/R_e)} (e^{\rho(1-R/R_e)} - 2), \quad (1)$$

where the potential between two particles, in terms of the distance between them,  $R$ , is parameterised by the pair equilibrium distance,  $R_e$ , by  $\epsilon$ , the pair well depth at  $R = R_e$ , and by the range parameter  $\rho$ . Taking  $\epsilon = 1$  and  $R_e = 1$  sets the energy and length units. The range relative to the particle diameter can be adjusted to represent different systems:  $\rho = 3$  is a long range, appropriate for sodium atoms,<sup>38</sup> while  $\rho = 14$  is a short range, suitable for  $C_{60}$ ;<sup>34</sup>  $\rho = 30$  is a very short range, which has been used for modelling colloids.<sup>39–43</sup>

The Morse potential is pairwise additive so the total energy of the system is a sum over the interaction between all pairs of particles. This approximation neglects many-body interactions, which can be important in some systems, especially when the particle interaction range is large.<sup>44</sup> The pair approximation has been successfully used to predict experimental properties, for colloids for example, when the range is small.<sup>39</sup>

In the present work we explore the PES for small atomic clusters bound by the Morse potential with a range of values for  $\rho$ , namely 3, 6, 10, 14 and 30. Our databases are almost complete, meaning that they were expanded until no new minima were found, but the search methods used do not rigorously guarantee that all the minima have been located. Disconnectivity graphs are used to visualise the potential energy landscape.<sup>45,46</sup> Previously, global minima and some low-lying minima have been described,<sup>31–37,42,47</sup> but attempts to generate relatively complete databases have been made only for a few specific cases.<sup>30,43</sup>

## 2 Methods

### 2.1 Generating Databases

Clusters in two and three dimensions bound by the Morse potential were analysed. For each number of particles  $N$  and range  $\rho$  a database of potential energy minima and the transition states connecting them was generated. First, the global minimum and other low-lying minima were located by basin-hopping global optimisation<sup>1,48–52</sup> as implemented in GMIN.<sup>53</sup>  $10^4$  basin-hopping steps were run from each of three random starting configurations. The twenty lowest minima from each run, or all minima if there were fewer than twenty, were then connected *via* transition states located using the doubly-nudged<sup>54–56</sup> elastic band<sup>57–61</sup> and eigenvector-following<sup>62</sup> algorithms in OPTIM.<sup>63</sup> These minima and transition states were imported into PATHSAMPLE,<sup>64</sup> at which point duplicate minima and transition states were removed. Additional single-ended transition state searches<sup>62,65,66</sup> and searches using the UNTRAP<sup>67</sup> procedure were performed to expand the database. After no more transition states were located, a further 50 single-ended searches were performed for each minimum. At this point we can be reasonably confident that practically all the minima and transition states have been found.

For each database, a network was constructed by considering the minima as nodes with edges between any two minima directly connected by a transition state. The barrier height for the transition state was ignored and the path through the transition state can be followed in either direction, so the network was treated as undirected and unweighted. The network was analysed using the Python NetworkX package.<sup>68</sup> A further check on the completeness of the database was to ensure that all minima were connected to each other by one or more transition states, *i.e.* that the network was connected. Finally, transition states connecting the same pair of minima as an earlier pathway were removed, to leave at most one transition state connecting any pair. Transition states that connected permutation-inversion isomers of the same minimum, which correspond to degenerate rearrangements,<sup>69</sup> were also removed.

We chose to study clusters with  $\rho = 3, 6, 10, 14$  and 30, in both two and three dimensions. These range parameters correspond to previous values used for the Morse potential and give a good spread of physically relevant pair potentials. The number of particles was chosen to produce databases with between several hundred and a few thousand minima. Selecting these sizes meant we could be confident of generating essentially complete databases in a reasonable time, improving the reliability of statistical tests. The symbol  $M_N^{x,D}$  is used to refer to a Morse cluster in  $x$  dimensions with  $N$  particles.

### 2.2 Degree Distribution

A node  $s$  has degree  $k_s$ , defined as the number of edges connected to  $s$ . The degree distribution  $p_k$  is the number of nodes with degree  $k$ , or after normalisation the fraction of nodes with degree  $k$ , or the probability of a selected node having degree  $k$ .<sup>5</sup>

We are interested in whether the degree distribution follows a power law,<sup>11</sup> such that

$$\begin{aligned} p_k &= Ck^{-\alpha}, \\ \ln p_k &= \ln C - \alpha \ln k, \end{aligned} \quad (2)$$

where  $C$  is a constant that is set by the normalisation, and  $\alpha$  is to be determined by the best fit to the data. Clauset *et al.*<sup>15</sup> have suggested a statistical test based on the method of maximum likelihood for determining the best fit. The maximum likelihood estimator (MLE) is:<sup>70</sup>

$$\alpha = 1 + n_k \left[ \sum_{k_i \geq k_{min}} \ln \frac{k_i}{k_{min} - \frac{1}{2}} \right]^{-1}, \quad (3)$$

where  $n_k$  here is the number of nodes with degree at least  $k_{min}$  and the sum runs over only those nodes. This formula gives an estimate of the best value of  $\alpha$ , but gives no indication of the quality of the fit.

Typically, a power law is only obeyed in the tail of the distribution, so a cutoff  $k_{min}$  above which the power law applies needs to be determined. Clauset *et al.* also have a method for determining the best value of  $k_{min}$ .<sup>71</sup> A fit is calculated for all possible values of  $k_{min}$  and the value which gives the smallest Kolmogorov-Smirnov (KS) statistic<sup>72</sup> is chosen.

A goodness-of-fit test is used to check whether the proposed power law is a plausible fit for the measured data.<sup>15</sup> Data sets are generated from the power law and compared to the measured data. The  $p$  value is the fraction of generated data sets that are a worse fit than the measured data. The fit is rejected if the  $p$  value is less than 0.05. If  $p > 0.05$ , it does not necessarily indicate that the power law is a good fit, merely that we do not have sufficient evidence to reject it.

Other distributions must be considered that could be a better fit to the data, such as the log-normal distribution. If we cannot say that a power law is a better fit than the log-normal, we should not conclude that the data follows a power law.<sup>15</sup> A likelihood ratio test is used to compare the fit of two distributions, indicating which distribution is more likely, and a  $p$  value is used to judge whether the result is significant.<sup>73</sup> A small  $p$  value ( $< 0.05$ ) means there is sufficient evidence to accept the result of the tests.

## 2.3 Clustering

Clustering in a network is the degree to which the neighbours of a node are also neighbours of each other. A local clustering coefficient can be defined for an individual node  $s$  as:<sup>6</sup>

$$c_s = \frac{\text{number of pairs of neighbours of } s \text{ that are connected}}{\text{number of pairs of neighbours of } s}. \quad (4)$$

To evaluate the numerator, for each of the nodes present in the adjacency list of  $s$ , we count how many of the adjacency lists of the other neighbours contain the same node. The denominator is easy to evaluate, as  $\frac{1}{2}k_s(k_s - 1)$ . The local clustering coefficient is an indicator of how important a node is in allowing flow through the network in its local neighbourhood.

The local clustering coefficient can be extended to a global clustering coefficient for the network in two ways. The average clustering coefficient is the mean of the local clustering coefficients:<sup>6</sup>

$$C_{av} = \frac{1}{n} \sum_s c_s. \quad (5)$$

The transitivity<sup>74</sup> depends on the number of triads and triangles in the network. A triad exists if nodes  $r$  and  $s$  are neighbours and nodes  $s$  and  $t$  are neighbours. A triad is a triangle if  $r$  and  $t$  are also neighbours. The transitivity is given by:

$$C_T = 3 \times \frac{\text{number of triangles}}{\text{number of triads}}, \quad (6)$$

where the factor of three accounts for the triple counting of triads that are also triangles. The transitivity can be calculated by extending the calculation of the local clustering coefficient to look at the neighbours of every node. The numerator is the number of pairs of neighbours of all nodes that are also connected. The denominator is  $\sum_s \frac{1}{2}k_s(k_s - 1)$ . The algorithmic complexity depends on the degree distribution, but in the worst case is  $O(n^2)$ .

These two measures differ in the weighting they give to nodes of small degree. Such nodes have a small number of possible neighbour pairs, so they have a small denominator in their local clustering coefficient. They therefore dominate the sum in equation (5). In networks with a large number of small degree nodes, the average clustering coefficient does not depend much on the larger degree nodes. In PES networks, the large degree nodes tend to be those with a low potential energy, which are considered more important for the properties of the network. We therefore prefer the transitivity to the average clustering coefficient.

To determine whether or not the transitivity should be considered large or small, it is useful to compare it to the value for an Erdős-Rényi random network<sup>75</sup> with the same number of nodes and edges. In this model, every edge has an equal probability of being present, independent of the existence of

any other edge. Therefore the probability of a given triad being a triangle, which is the probability of the final edge being present, is the probability of any edge existing. The maximum number of edges is the number of possible pairs of nodes,  $n(n - 1)/2$ , so given the number of edges ( $m$ ) and nodes ( $n$ ), the transitivity is:

$$C_{Random} = \frac{2m}{n(n - 1)}. \quad (7)$$

## 2.4 Average Shortest Path Length

A path in a network is a sequence of steps between nodes along edges. For any pair of nodes the shortest path (or geodesic path) is defined as the one connecting the two nodes containing the least number of edges.<sup>5</sup> The shortest path need not be unique, but here we are concerned only with its length, which is the number of edges it contains. If no path exists between a pair of nodes, then the network is unconnected. The complete network for a PES must be connected, as it is impossible to have minima on the surface that cannot be reached from any other minima through transition states for models with continuous degrees of freedom. The length of the shortest path between one node  $s$  and all others can be calculated using a breadth-first search.<sup>76</sup>

## 2.5 Assortativity

The assortativity<sup>77</sup> of a network is a measure of the tendency of nodes to have connections with other similar nodes. Specifically, the degree assortativity refers to whether nodes connect to other nodes of a similar degree. The assortativity coefficient takes values between 1 and  $-1$ , where 1 indicates connections between nodes of similar degree, 0 indicates no preference and  $-1$  means that high degree nodes have a preference to connect to low degree nodes. The definition is,

$$\rho_D = \frac{\sum_{jk} jk (e_{jk} - q_j q_k)}{\sigma_q^2}, \quad (8)$$

where  $j$  and  $k$  are the remaining degrees of two nodes (the degree of the node excluding the edge connecting the pair),  $q_j$  is the remaining degree distribution of  $j$ ,  $e_{jk}$  is the joint probability distribution of  $q_j$  and  $q_k$ , and  $\sigma_q$  is the standard deviation of  $q$ . The sum runs over the possible values of  $j$  and  $k$ . This is equivalent to the Pearson correlation coefficient<sup>78</sup> between degrees of nodes at the ends of each edge.

## 3 Results

Table 1 shows the number of minima and transition states found for the largest network considered at each range in two



Dimensions	$\rho$	$N$	Minima	Transition States
2	3	27	1135	17006
2	6	14	805	13098
2	10	14	840	19965
2	14	14	843	20552
2	30	13	358	6966
3	3	30	663	6678
3	6	13	1478	25173
3	10	12	2258	41441
3	14	12	2980	60615
3	30	11	1127	13752

**Table 1** The sizes of the largest networks considered for Morse clusters of  $N$  atoms in 2 and 3 dimensions.

and three dimensions, after the removal of transition states corresponding to alternative connections and degenerate rearrangements. Figure 1 illustrates the number of minima and transition states for all the clusters considered.

Some general patterns are clear: 3D clusters have more minima for the same range and number of particles than 2D clusters, except at very long-range with  $\rho = 3$ . The well-known approximate exponential growth of the number of minima and transition states with system size is present.<sup>4,79</sup> A shorter-range potential produces an energy landscape that is locally rough, having more minima, but globally smooth with a less overall funnelled structure.<sup>42</sup> To avoid the databases growing too large to be confident of finding the complete network, it was necessary to restrict the number of particles in the cluster as the range decreased. The difference in the database size is more apparent in three dimensions than two dimensions. This effect is evident in the disconnectivity graphs in figure 2, where the funnel for a  $\rho = 6$  landscape is apparent, suggesting structure-seeking behaviour, but there is no such structure for  $\rho = 30$ , where the landscape is more frustrated in the region of the global minimum.

### 3.1 Small-World Behaviour

To test whether or not the networks show small-world behaviour, the average shortest path length and the transitivity were considered. The average shortest path lengths for all the clusters are given in figure 3, and they show the expected logarithmic dependence on the number of nodes. The path lengths become slightly smaller as the range decreases. The number of transition states increases more rapidly with decreasing range than the number of minima, so a small decrease in the path length with decreasing range is not surprising.

The transitivity for a small-world network is expected to be larger than for the equivalent random graph with the same number of minima and edges.<sup>6</sup> The transitivity scaled by the

value for a random graph with the same number of nodes and edges is plotted for all the clusters in figure 4. The expected small-world behaviour is again apparent, with the transitivity larger than for a random graph in all the clusters, except for three-dimensional systems with very few minima, where finite size effects dominate. There is no significant difference between the values for different ranges, except for  $\rho = 3$ , which appears to have a slightly higher transitivity. The long range of the potential produces a smoother landscape, which makes it more likely that minima in distant parts of configuration space are directly connected. While short-range clusters do not display the structure-seeking properties of long-range clusters, since their landscapes are not funnelled, this is due to the presence of large barriers. Since the present analysis accounts only for the existence of connections, regardless of barrier heights, we find that the small-world properties of the network are similar for short- and long-range clusters.

Insight into the structure of the landscape and network can be obtained by inspecting a plot of the degree of the nodes against their potential energy, shown in figure 5 for  $M_{13}^{3D}$  with  $\rho = 6$  and the  $M_{11}^{3D}$  with  $\rho = 30$ . The funnelled landscape for the  $\rho = 6$  cluster has one node with a significantly lower potential energy, as seen in the disconnectivity graph (figure 2), which is also the node with the highest degree. Over a third of the minima in the network are directly connected to the global minimum. At higher energies, although there are many minima with different degrees at similar energies, the maximum degree for a given energy decreases as the energy increases. These observations can be related to funnels on the landscape: the global minimum has a large basin of attraction, which occupies a significant volume in configuration space, giving a large boundary surface containing many transition states. High energy minima have small basins of attraction with a small boundary surface and fewer transition states. In comparison, the global minimum for  $M_{11}^{3D}$  with  $\rho = 30$  is not the minimum with the highest degree, and the most highly connected minimum has transition states linking it to less than a quarter of the other minima. The general trend for higher energy minima to have a lower degree is still present. The prominent vertical stripes are a result of the short-ranged potential. Since the potential energy is primarily determined by the number of nearest-neighbour contacts, an integer, it is approximately discretised.<sup>80–84</sup>

### 3.2 Scale-Free Behaviour

To test whether a network is scale-free, we must determine whether the degree distribution (or cumulative degree distribution) follows a power law, at least in the high degree range. The fitted  $\alpha$  and  $k_{min}$  parameters [see equation (2)] and  $p$  values from the goodness-of-fit tests for the largest cluster for each dimension and range parameter are shown in table 2. We

Dimensions	$\rho$	N	$\alpha$	$k_{min}$	$n_{tail}$	$p$
2	3	27	$2.32 \pm 0.32$	$23 \pm 12$	255	0.00
2	6	14	$3.50 \pm 0.16$	$26 \pm 4$	302	0.00
2	10	14	$2.55 \pm 0.14$	$18 \pm 7$	765	0.00
2	14	14	$2.03 \pm 0.21$	$10 \pm 5$	840	0.00
2	30	13	$2.90 \pm 0.15$	$17 \pm 4$	332	0.00
3	3	30	$2.79 \pm 0.17$	$10 \pm 2$	219	0.18
3	6	13	$2.66 \pm 0.16$	$19 \pm 8$	698	0.02
3	10	12	$2.56 \pm 0.09$	$28 \pm 3$	498	0.03
3	14	12	$2.37 \pm 0.12$	$65 \pm 11$	174	0.02
3	30	11	$3.30 \pm 0.25$	$48 \pm 8$	47	0.16
3	LJ	14	$2.82 \pm 0.10$	$37 \pm 8$	513	0.07

**Table 2** Fitted values of  $\alpha$ ,  $k_{min}$  and  $p$  values from the goodness-of-fit tests for the largest Morse cluster considered at each range and dimension, as well as LJ<sub>14</sub>. If  $p < 0.05$ , the fit should be considered poor.

also show  $n_{tail}$ , the number of minima remaining after those with degree less than  $k_{min}$  have been removed. If  $p < 0.05$ , there is good evidence to reject the hypothesis that a power law is a good fit to the data; otherwise the power law may be a good fit. A large value of  $k_{min}$  and a small value of  $n_{tail}$  indicates the fit is only applicable for a small range of the data and is unreliable.

The cumulative degree distribution is plotted in figure 6 for  $M_{13}^{3D}$  clusters with  $\rho = 6$  (top) and  $M_{11}^{3D}$  with  $\rho = 30$  (bottom), along with the best power law fit calculated by the MLE method. The low degree end does not follow a power law, because the number of minima for a cluster is finite. Above  $k_{min}$ , the distribution appears to be approximately linear. However, the goodness-of-fit test for the  $\rho = 6$  cluster gives  $p < 0.05$ , indicating the power law is not a good fit for the data.

In fact, according to the suggested criteria, only two clusters have a  $p$  value compatible of a power law fit:  $M_{30}^{3D}$  with  $\rho = 3$  and  $M_{11}^{3D}$  with  $\rho = 30$ . Both of these clusters have a large value of  $k_{min}$  compared to the number of minima in the cluster, giving a small value of  $n_{tail}$ . Therefore the fit should be considered unreliable, especially as it is clearly the case that these clusters do not generally display scale-free behaviour.

### 3.3 Lennard-Jones Clusters

Doye and Massen<sup>17</sup> studied Lennard-Jones (LJ) clusters and concluded that the degree distribution follows a power law and that the networks show scale-free behaviour. However, they did not give a statistical analysis, as the appropriate tools had not been identified at that time, simply stating that ‘as the size of the clusters increase a clear power law tail develops.’ Using their data for the cluster with 14 particles,<sup>85</sup> which includes 4196 minima and 61085 transition states, we have now applied the goodness-of-fit tests, as for the Morse clusters. These data

are shown in table 2 and the cumulative degree distribution and fit are shown in figure 7. Since  $p > 0.05$ , this distribution is a candidate for a power law fit. However, we tested the power law fit against a log-normal fit. The log-likelihood ratio  $R = 0.80$ , but  $p = 0.41$ , so the test is inconclusive. Clauset *et al.* comment that deciding between a power law fit and a log-normal fit can be problematic.<sup>15</sup> We conclude that the distribution approximately follows a power law.

### 3.4 Bulk Ortho-Terphenyl

A database of minima and transition states for a bulk representation of the molecular glass-former ortho-terphenyl (OTP) was provided by Niblett and coworkers.<sup>26</sup> The details of the model are described in their paper. The database was relatively large, containing 313,651 minima and 334,272 transition states. After removing alternative transition states linked the same pairs of minima and self-connections (degenerate rearrangements), 332,774 transition states remained. In contrast to the Morse and Lennard-Jones clusters considered above, the database is certainly very far from complete, since finding all the minima and transition states for such a glassy landscape is unfeasible.

Testing whether the incomplete OTP database is likely to be representative of the full network was assessed by taking increasingly large subsets of the database and looking at the convergence of the network properties. Here the appropriate network properties are the ones that are global, but do not directly depend on the size of the network. The average shortest path length is therefore unsuitable, as it is expected to increase with the network size. Three properties were chosen: the transitivity, the average clustering coefficient (the mean of the local clustering coefficients) and the assortativity.

All three properties are only defined for a connected network, but an arbitrary subset of the database will not generally be connected because of the way it was constructed. Hence transition states were added to the network one by one, in the order located by the original search, and the largest connected component selected. Whenever the largest connected component increased in size, the network properties were calculated. Graphs of the three properties against the number of transition states in the connected component are shown in figure 8. There are fluctuations in the values, but all three seem to be approaching a limit. However, caution is needed in interpreting these results. Sampling biases in generating the database could potentially lead to limiting values of these properties without truly being a representative sample of the network.

To test for small-world properties, the values of the average shortest path length and transitivity can be compared to those of a random graph. These ratios are plotted in figure 9 for increasingly large connected subgraphs of the network. The transitivity and average shortest path length increase more

quickly than the values for the equivalent random graph suggesting that the network (or at least this subset of the full PES) exhibits lattice graph properties, rather than small-world behaviour. The network is mildly assortative, although the value is close to zero, indicating a slight preference for minima to connect to other minima with similar degree. Assortative networks tend to have cores of highly connected nodes with a periphery of low degree nodes.<sup>5</sup> This pattern could arise in a glassy landscape, with well connected groups of structurally closely related minima, and higher energy minima forming the periphery between the numerous groups.

Scale-free networks are small-world, so this network is also not scale-free. We found  $p = 0.0$  for the power law goodness-of-fit test.

### 3.5 Bulk Binary Lennard-Jones

A database for a bulk BLJ system was provided by de Souza and Wales.<sup>27,28</sup> The system contained 60 atoms, 48 of type A and 12 of type B, with interaction parameters  $\sigma_{AA} = 1.0$ ,  $\sigma_{AB} = 0.8$ ,  $\sigma_{BB} = 0.88$ ,  $\epsilon_{AA} = 1.0$ ,  $\epsilon_{AB} = 1.5$  and  $\epsilon_{BB} = 0.5$ . These parameters are known to give rise to a glassy landscape.<sup>86,87</sup> Periodic boundary conditions were used with a cubic box of length 3.587. The database was connected and contained 11,538 minima and 13,088 transition states, dropping to 12,957 transition states after removing alternative transition states linking the same pairs of minima and degenerate rearrangements.

The database was expanded until it contained 703,827 minima and 360,836 transition states. However, the largest connected component consisted of only 12,905 minima and 14,363 transition states.

Our analysis followed the steps described above for OTP. The convergence of the average clustering coefficient, transitivity and assortativity are shown in figure 10, and the ratios of the average shortest path length and transitivity to those for the equivalent random graph are shown in figure 11. The results are qualitatively similar to OTP, although the convergence of the average clustering coefficient and transitivity is less clear due to the smaller sample size. The value of the assortativity is negative, indicating slight disassortativity compared to the slight assortativity of the OTP system. Park and Newman suggest that disassortativity can arise due to the exclusion of alternative and self-connections,<sup>88</sup> so this result is unsurprising, although it is unclear why the opposite behaviour to OTP is observed. The growing ratios of average shortest path length and transitivity to the equivalent random graph suggest the same conclusion: that the system is showing lattice graph behaviour rather than small-world behaviour.

## 4 Conclusions

Networks of minima and transition states for Morse clusters with different ranges in the interatomic potential were analysed. The networks exhibited small-world properties, but statistical testing using recent methods demonstrated that they are probably not scale-free. The LJ<sub>14</sub> cluster was re-evaluated using these new statistical methods and the tests were inconclusive, and we cannot firmly conclude the network is scale-free, but it remains a possibility.

The Morse networks were compared with glassy systems, which have very different energy landscapes characterised by numerous low energy amorphous structures separated by relatively high barriers. The two glassy networks did not show small-world behaviour, which suggests that the small-world properties of the LJ and Morse networks are a feature of their single-funnel potential energy landscapes. However, further investigation is required. Larger clusters with more minima and transition states will make statistical tests for scale-free behaviour more reliable.

Using the convergence of network properties to test whether a sample is representative opens the possibility of studying systems where it is impractical to find the complete network. While it remains computationally intractable to find the complete network for glassy systems, as the number of minima and transition states is far too large, we believe the approach here will allow further insight for such systems. It is certainly possible to generate and analyse larger samples, and if databases generated by different schemes converge on the same properties we can have greater confidence about whether the samples are truly representative. Analysing the network properties of a representative sample will provide deeper understanding of the behaviour of bulk glasses. The LJ<sub>38</sub> cluster, which has a double-funnel landscape,<sup>89</sup> could produce interesting results if the two funnels are considered separately and together. Studying a fast-folding protein network would also be insightful. On the basis of the results here, we anticipate that the folding network is not really scale-free, but due to its known structure-seeking nature, small-world character is likely.

### Conflicts of Interest

There are no conflicts of interest to declare.

### Acknowledgements

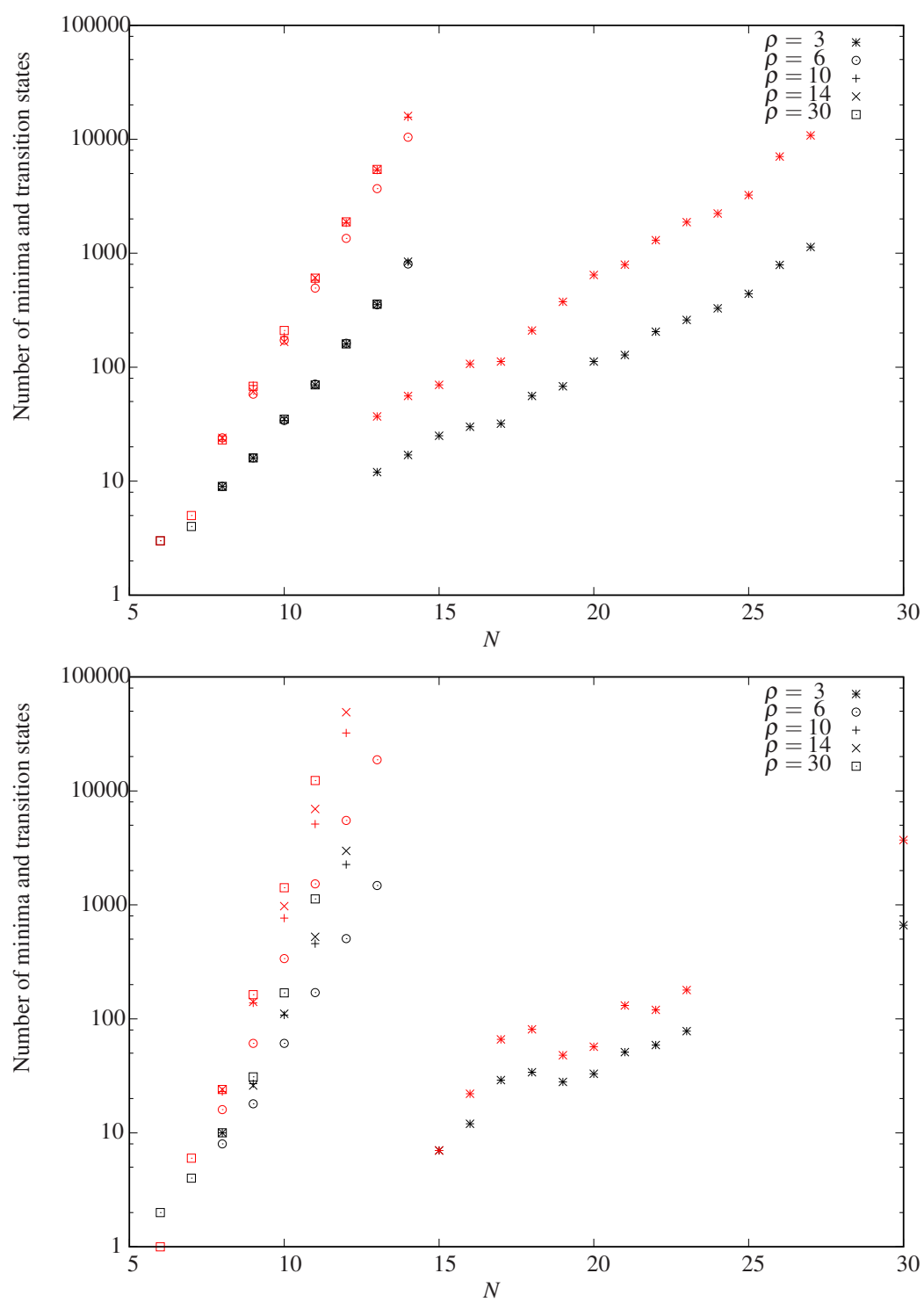
JM acknowledges the support of a Sackler Studentship from the University of Cambridge. A part of this work was done while DM was employed by the Department of Chemistry at the University of Cambridge. This work was financially supported by the Engineering and Physical Sciences Research Council and the European Research Council. Data can be accessed at <http://doi.org/10.5281/zenodo.580085>.

## References

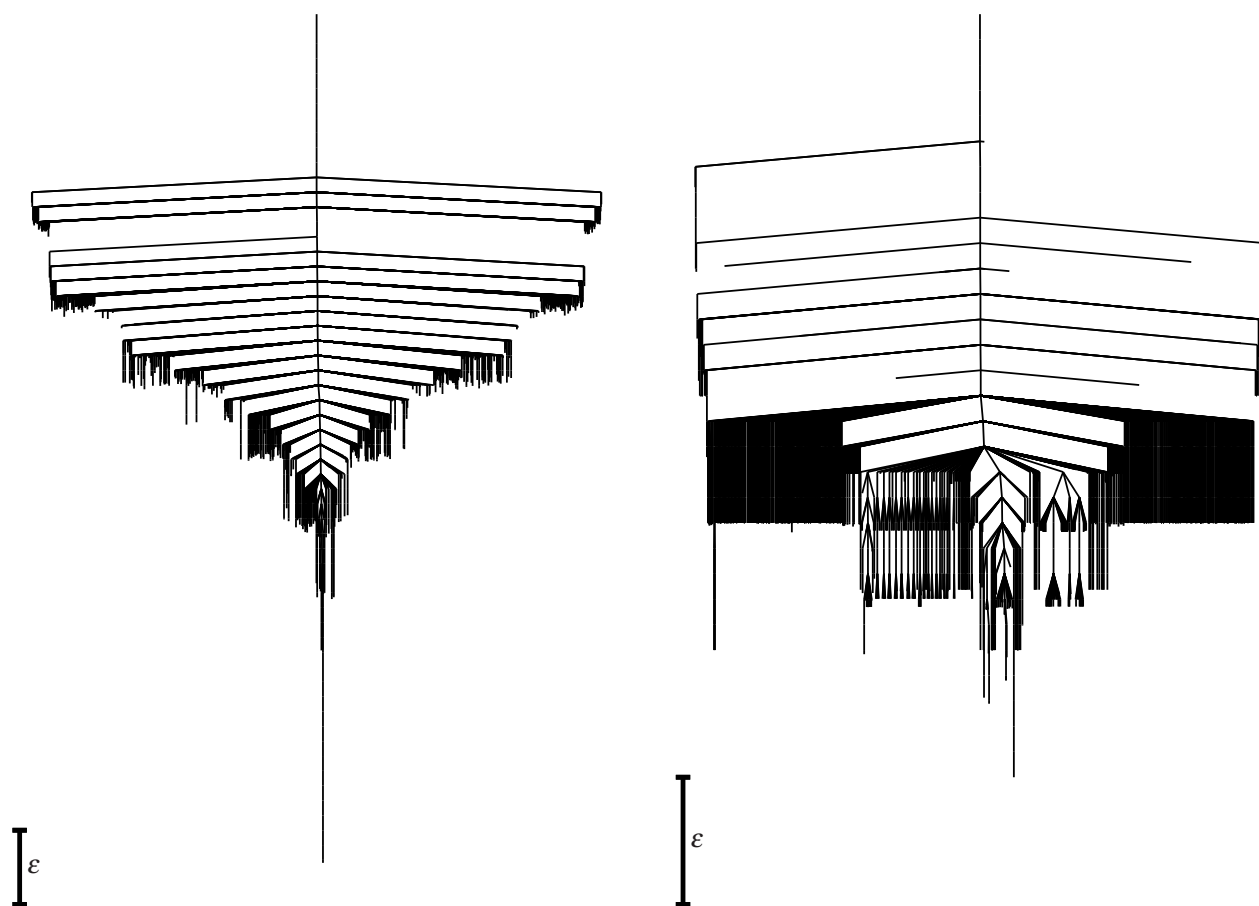
- 1 D. J. Wales, *Energy landscapes*, Cambridge University Press, 2003.
- 2 J. N. Murrell and K. J. Laidler, *Trans. Faraday Soc.*, 1968, **64**, 371–377.
- 3 F. H. Stillinger and T. A. Weber, *Phys. Rev. A*, 1982, **25**, 978–989.
- 4 F. H. Stillinger and T. A. Weber, *Science*, 1984, **225**, 983–989.
- 5 M. E. J. Newman, *Networks: An Introduction*, Oxford University Press, 2010.
- 6 D. J. Watts and S. H. Strogatz, *Nature*, 1998, **393**, 440–442.
- 7 A. Fronczak, P. Fronczak and J. A. Holyst, *Phys. Rev. E*, 2002, **70**, 056110.
- 8 S. Milgram, *Psychol. Today*, 1967, **2**, 6067.
- 9 J. Guare, *Six Degrees of Separation: A Play*, Vintage Books, New York, 1990.
- 10 The Oracle of Bacon, <http://oracleofbacon.org/>.
- 11 A.-L. Barabási and R. Albert, *Science*, 1999, **286**, 509–512.
- 12 R. Albert and A.-L. Barabási, *Rev. Mod. Phys.*, 2002, **74**, 47–97.
- 13 D. Hughes, M. Paczuski, R. O. Dendy, P. Helander and K. G. McClements, *Phys. Rev. Lett.*, 2003, **90**, 131101.
- 14 M. Baiesi and M. Paczuski, *Phys. Rev. E*, 2004, **69**, 066106.
- 15 A. Clauset, C. R. Shalizi and M. E. J. Newman, *SIAM Review*, 2009, **51**, 661–703.
- 16 J. P. K. Doye, *Phys. Rev. Lett.*, 2002, **88**, 238701.
- 17 J. P. K. Doye and C. P. Massen, *J. Chem. Phys.*, 2005, **122**, 084105.
- 18 J. M. Carr and D. J. Wales, *J. Phys. Chem. B*, 2008, **112**, 8760–8769.
- 19 J. P. K. Doye and C. P. Massen, *Phys. Rev. E*, 2005, **71**, 016128.
- 20 F. Rao and A. Caffisch, *J. Mol. Biol.*, 2004, **342**, 299–306.
- 21 G. R. Bowman and V. S. Pande, *Proc. Natl. Acad. Sci. USA*, 2010, **107**, 10890–10895.
- 22 D. Chakraborty, R. Collepardo-Guevara and D. J. Wales, *J. Am. Chem. Soc.*, 2014, **136**, 18052–18061.
- 23 D. Mehta, J. Chen, D. Z. Chen, H. Kusumaatmaja and D. J. Wales, *Phys. Rev. Lett.*, 2016, **117**, 028301.
- 24 J. J. Thomson, *Lond. Edinb. Dubl. Phil. Mag.*, 1904, **7**, 237–265.
- 25 A. J. Ballard, R. Das, S. Martiniani, D. Mehta, L. Sagun, J. D. Stevenson and D. J. Wales, *Phys. Chem. Chem. Phys.*, 2017, **19**, 12585–12603.
- 26 S. P. Niblett, V. K. de Souza, J. D. Stevenson and D. J. Wales, *J. Chem. Phys.*, 2016, **145**, 024505.
- 27 V. K. de Souza and D. J. Wales, *Phys. Rev. B*, 2006, **74**, 134202.
- 28 V. K. de Souza and D. J. Wales, *J. Chem. Phys.*, 2008, **129**, 164507.
- 29 P. M. Morse, *Phys. Rev.*, 1929, **34**, 57–64.
- 30 P. A. Braier, R. S. Berry and D. J. Wales, *J. Chem. Phys.*, 1990, **93**, 8745–8756.
- 31 J. P. K. Doye, D. J. Wales and R. S. Berry, *J. Chem. Phys.*, 1995, **103**, 4234–4249.
- 32 D. T. Mainz and R. S. Berry, *Mol. Phys.*, 1996, **88**, 709–726.
- 33 J. P. K. Doye and D. J. Wales, *Science*, 1996, **271**, 484–487.
- 34 J. P. K. Doye and D. J. Wales, *J. Phys. B*, 1996, **29**, 4859–4894.
- 35 J. P. K. Doye and D. J. Wales, *J. Chem. Soc. Faraday T.*, 1997, **93**, 4233–4243.
- 36 L. Cheng and J. Yang, *J. Phys. Chem. A*, 2007, **111**, 5287–5293.
- 37 Y. Feng, L. Cheng and H. Liu, *J. Phys. Chem. A*, 2009, **113**, 13651–13655.
- 38 L. A. Girifalco and V. G. Weizer, *Phys. Rev.*, 1959, **114**, 687–690.
- 39 D. J. Wales, *ChemPhysChem*, 2010, **11**, 2491–2494.
- 40 A. Malins, S. R. Williams, J. Eggers, H. Tanaka and C. P. Royall, *J. Phys. Condens. Mat.*, 2009, **21**, 425103.
- 41 J. Taffs, A. Malins, S. R. Williams and C. P. Royall, *J. Phys. Condens. Mat.*, 2010, **22**, 104119.
- 42 F. Calvo, J. P. K. Doye and D. J. Wales, *Nanoscale*, 2012, **4**, 1085–1100.
- 43 J. W. R. Morgan and D. J. Wales, *Nanoscale*, 2014, **6**, 10717–10726.
- 44 J. W. Merrill, S. K. Sainis and E. R. Dufresne, *Phys. Rev. Lett.*, 2009, **103**, 138301.
- 45 O. M. Becker and M. Karplus, *J. Chem. Phys.*, 1997, **106**, 1495–1517.
- 46 D. J. Wales, M. A. Miller and T. A. Walsh, *Nature*, 1998, **394**, 758–760.
- 47 L. Cheng and J. Yang, *J. Phys. Chem. A*, 2007, **111**, 2336–2342.
- 48 Z. Li and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 6611–6615.
- 49 Z. Li and H. A. Scheraga, *J. Mol. Struct-Theochem*, 1988, **179**, 333.
- 50 D. J. Wales and J. P. K. Doye, *J. Phys. Chem. A*, 1997, **101**, 5111–5116.
- 51 D. J. Wales and H. A. Scheraga, *Science*, 1999, **285**, 1368–1372.
- 52 D. J. Wales and T. V. Bogdan, *J. Phys. Chem. B*, 2006, **110**, 2076520776.
- 53 D. J. Wales, GMIN: A program for basin-hopping global optimisation, <http://www-wales.ch.cam.ac.uk/software.html>.
- 54 S. A. Trygubenko and D. J. Wales, *J. Chem. Phys.*, 2004, **120**, 2082–2094.
- 55 S. A. Trygubenko and D. J. Wales, *J. Chem. Phys.*, 2004, **120**, 7820–7820.
- 56 D. Sheppard, R. Terrell and G. Henkelman, *J. Chem. Phys.*, 2008, **128**, 134106.
- 57 G. Mills and H. Jónsson, *Phys. Rev. Lett.*, 1994, **72**, 1124–1127.
- 58 G. Mills, H. Jónsson and G. K. Schenter, *Surf. Sci.*, 1995, **324**, 305–337.
- 59 H. Jónsson, G. Mills and K. W. Jacobsen, *Classical and Quantum Dynamics in Condensed Phase Simulations*, World Scientific, Singapore, 1998.
- 60 G. Henkelman, B. P. Uberuaga and H. Jónsson, *J. Chem. Phys.*, 2000, **113**, 9901–9904.
- 61 G. Henkelman and H. Jónsson, *The Journal of Chemical Physics*, 2000, **113**, 9978–9985.
- 62 L. J. Munro and D. J. Wales, *Phys. Rev. B*, 1999, **59**, 3969–3980.
- 63 D. J. Wales, OPTIM: A Program for Optimizing Geometries and Calculating Reaction Pathways, <http://www-wales.ch.cam.ac.uk/software.html>.
- 64 D. J. Wales, PATHSAMPLE: A program for generating connected stationary point databases and extracting global kinetics, <http://www-wales.ch.cam.ac.uk/software.html>.
- 65 D. J. Wales, J. P. K. Doye, M. A. Miller, P. N. Mortenson and T. R. Walsh, *Adv. Chem. Phys.*, 2000, **115**, 1.
- 66 Y. Kumeda, D. J. Wales and L. J. Munro, *Chem. Phys. Lett.*, 2001, **341**, 185–194.
- 67 B. Strodel, C. S. Whittleston and D. J. Wales, *J. Am. Chem. Soc.*, 2007, **129**, 16005–16014.
- 68 A. A. Hagberg, D. A. Schult and P. J. Swart, Proceedings of the 7th Python in Science Conference (SciPy2008), Pasadena, CA USA, 2008, pp. 11–15.
- 69 R. E. Leone and P. von R. Schleyer, *Angew. Chem. Int. Edit.*, 1970, **9**, 860–890.
- 70 A. N. M. Muniruzzaman, *Calcutta Statist. Assoc. Bull.*, 1957, **7**, 115–126.
- 71 A. Clauset, M. Young and K. S. Gleditsch, *J. Confl. Resolut.*, 2007, **51**, 58–87.
- 72 W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, 2nd edn., 1992.
- 73 Q. H. Vuong, *Econometrica*, 1989, **57**, 307–333.
- 74 R. D. Luce and A. D. Perry, *Psychometrika*, 1949, **14**, 95–116.
- 75 P. Erdős and A. Rényi, *Publ. Math. Debrecen*, 1959, **6**, 290–297.
- 76 E. F. Moore, *The shortest path through a maze*, Bell Telephone System., 1959.
- 77 M. E. J. Newman, *Phys. Rev. Lett.*, 2002, **89**, 208701.
- 78 K. Pearson, *Proc. R. Soc.*, 1895, **58**, 240–242.
- 79 D. J. Wales and J. P. K. Doye, *J. Chem. Phys.*, 2003, **119**, 12409–12416.
- 80 D. J. Wales, *J. Chem. Soc. Faraday T.*, 1994, **90**, 1061–1063.
- 81 J. P. K. Doye and D. J. Wales, *Chem. Phys. Lett.*, 1996, **262**, 167–174.
- 82 J. P. K. Doye, A. Dullweber and D. J. Wales, *Chem. Phys. Lett.*, 1997, **269**, 408–412.
- 83 W. Branz, N. Malinowski, A. Enders and T. P. Martin, *Phys. Rev. B*, 2002, **66**, 094107.



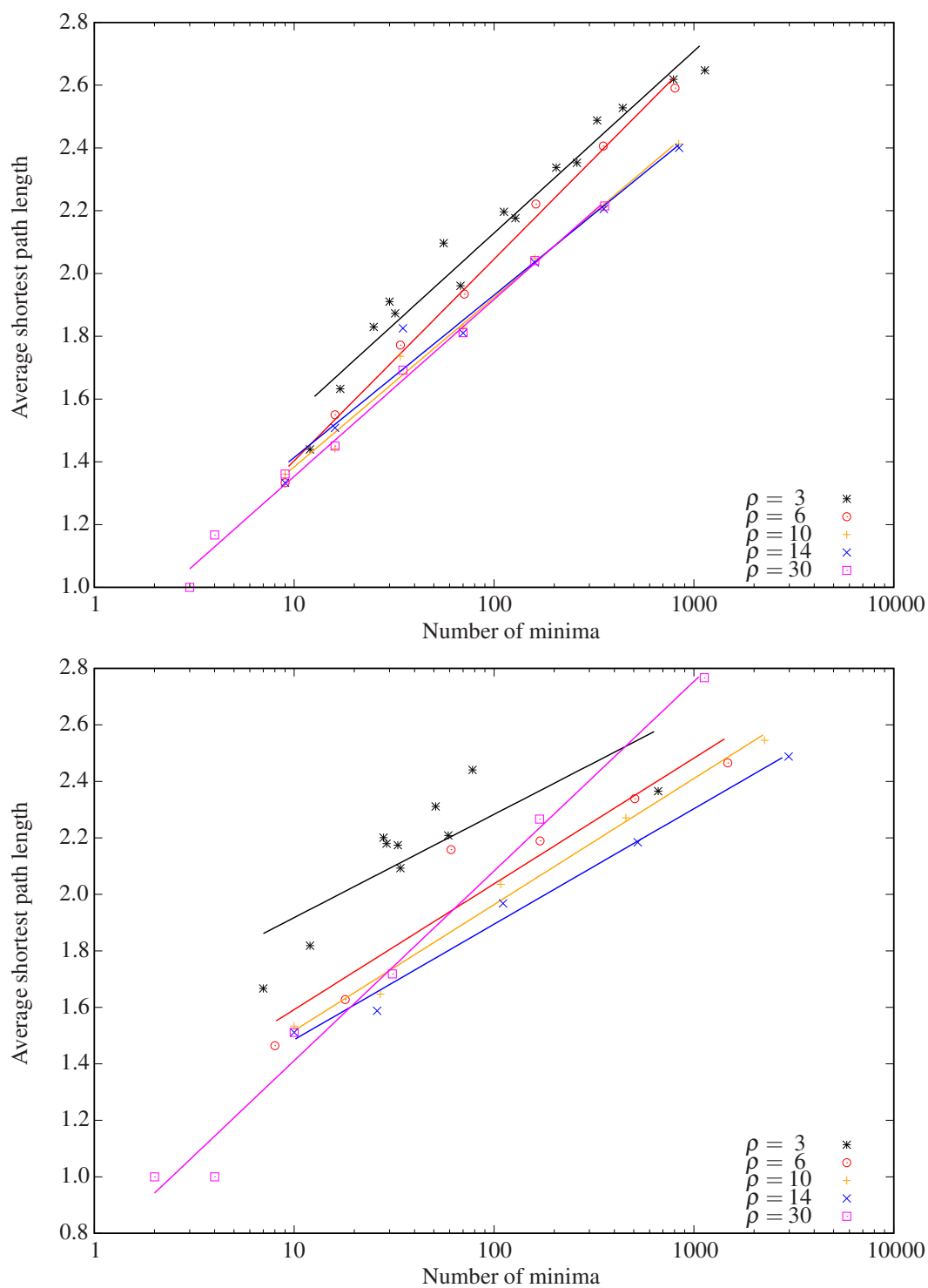
- 
- 84 F. Calvo, M. Benali, V. Gerbaud and M. Hemati, *Computing Lett.*, 2005, **1**, 183–191.
- 85 LJ<sub>N</sub> networks, <http://doye.chem.ox.ac.uk/networks/LJn.html>.
- 86 W. Kob and H. C. Andersen, *Phys. Rev. Lett.*, 1994, **73**, 1376–1379.
- 87 W. Kob and H. C. Andersen, *Phys. Rev. E*, 1995, **51**, 4626–4641.
- 88 J. Park and M. E. J. Newman, *Phys. Rev. E*, 2003, **68**, 026112.
- 89 J. P. K. Doye, M. A. Miller and D. J. Wales, *J. Chem. Phys.*, 1999, **110**, 6896–6906.



**Fig. 1** The number of minima (black) and transition states (red) for all the two-dimensional (top) and three-dimensional (bottom) clusters considered.

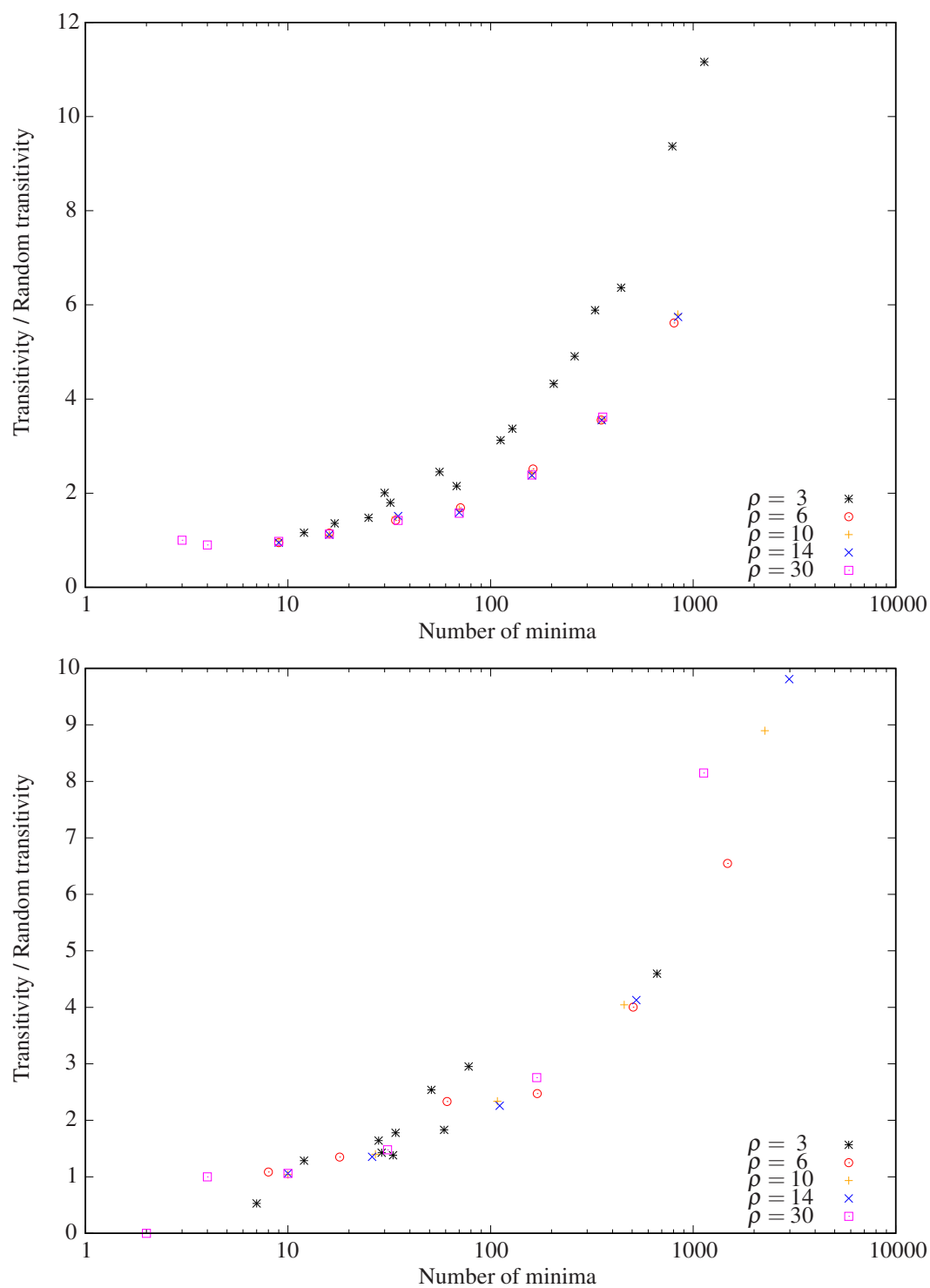


**Fig. 2** Disconnectivity graphs for  $M_{13}^{3D}$  with  $\rho = 6$  (left) and  $M_{11}^{3D}$  with  $\rho = 30$  (right). The difference in the global structure of the landscape is clear.

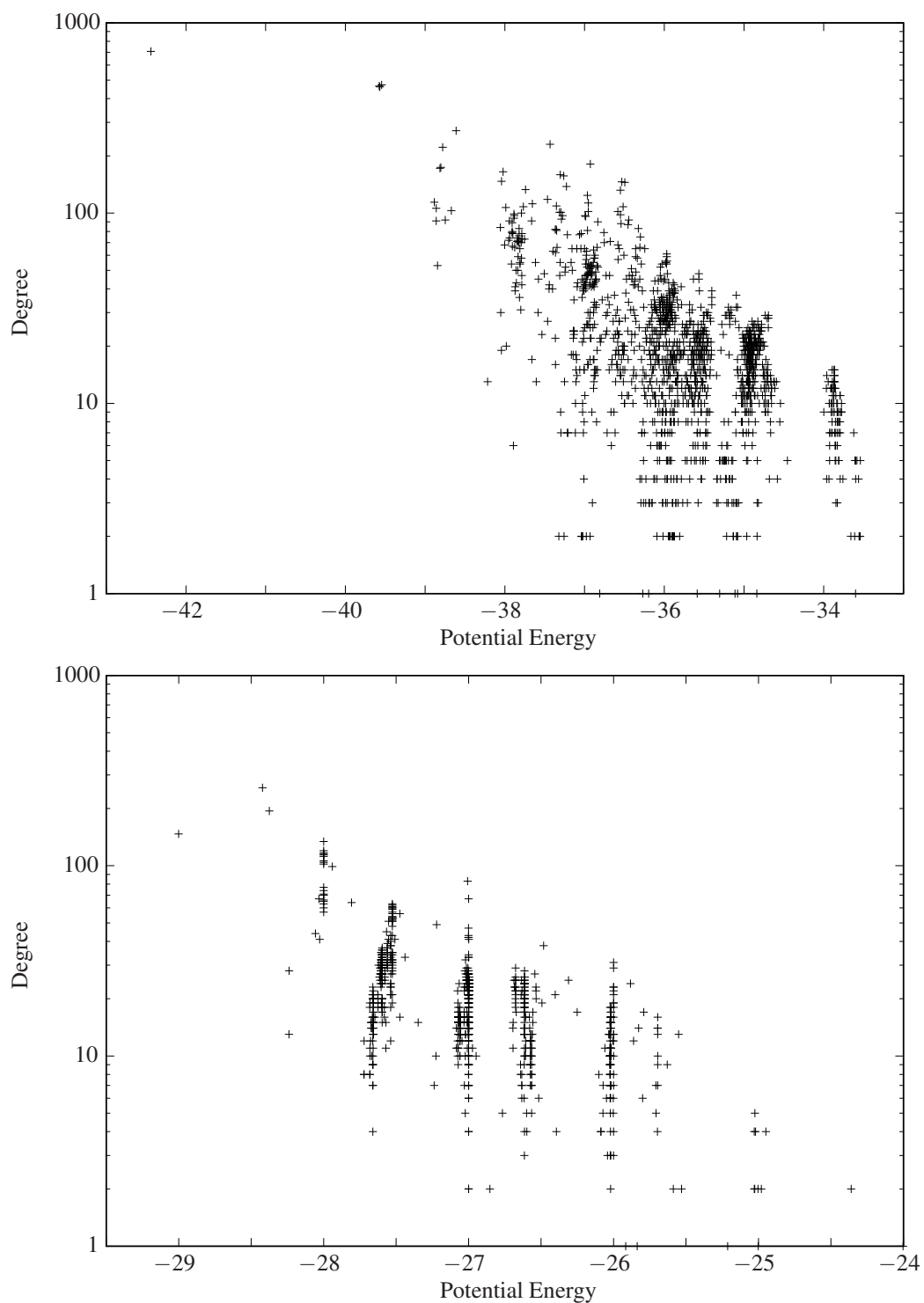


**Fig. 3** Plots of the average shortest path length against the number of minima for all the 2D clusters (top) and all the 3D clusters (bottom). Note the logarithmic scale on the horizontal axis.

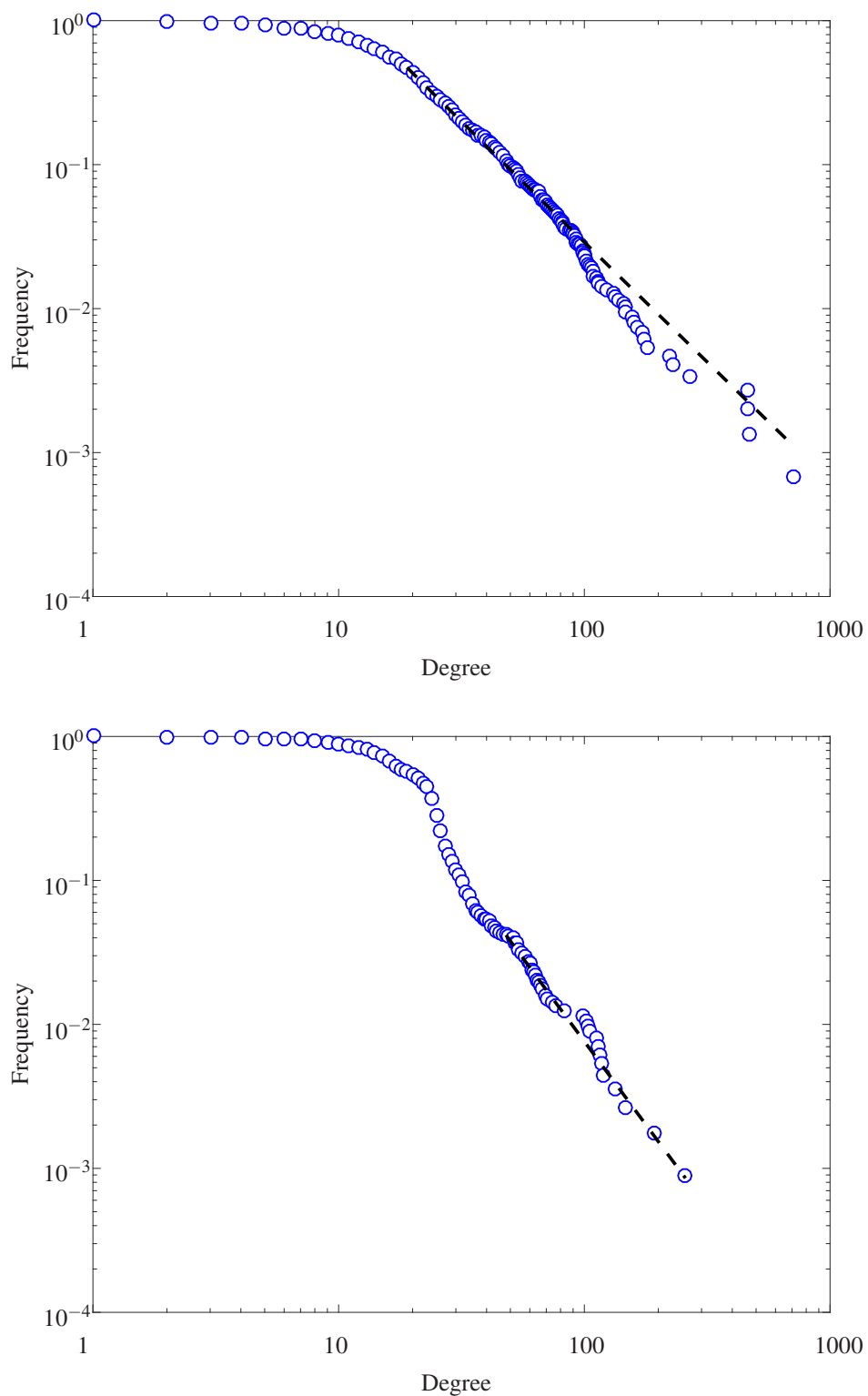




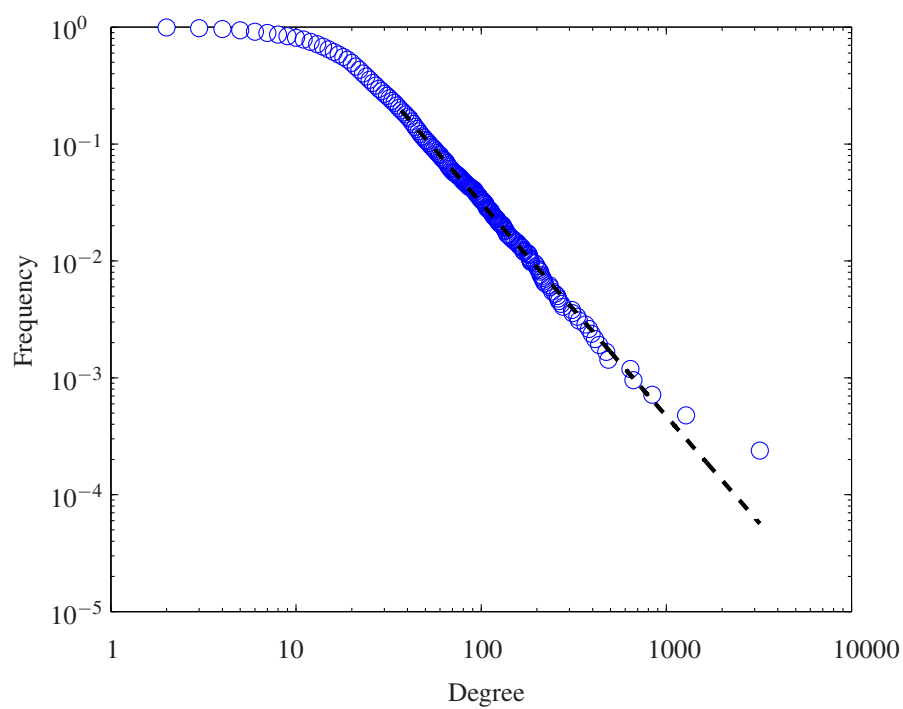
**Fig. 4** Plots of the ratio of the transitivity of each network to the transitivity of a random graph with the same number of nodes and edges, against the number of minima for all the 2D clusters (top) and all the 3D clusters (bottom).



**Fig. 5** Degree of each minimum as a function of potential energy for  $M_{13}^{3D}$  with  $\rho = 6$  (top) and  $M_{11}^{3D}$  with  $\rho = 30$  (bottom)

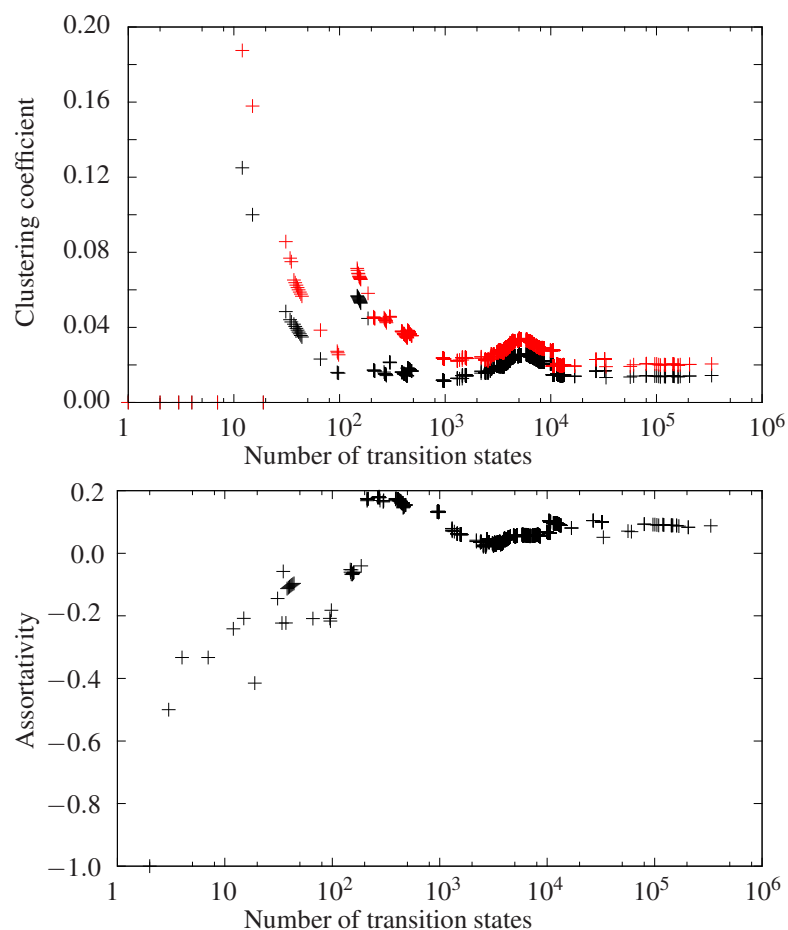


**Fig. 6** Plots of the normalised cumulative degree distribution for  $M_{13}^{3D}$  with  $\rho = 6$  (top) and  $M_{11}^{3D}$  with  $\rho = 30$  (bottom). The calculated best fit is indicated by the black dashed line.

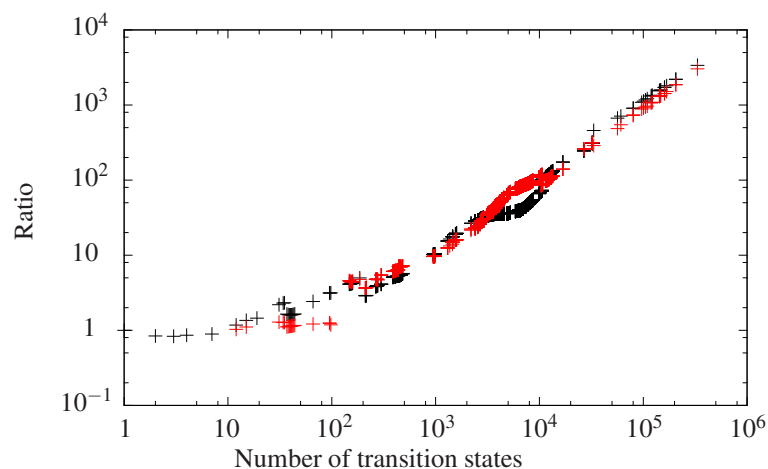


**Fig. 7** Plot of the normalised cumulative degree distribution for LJ<sub>14</sub>.

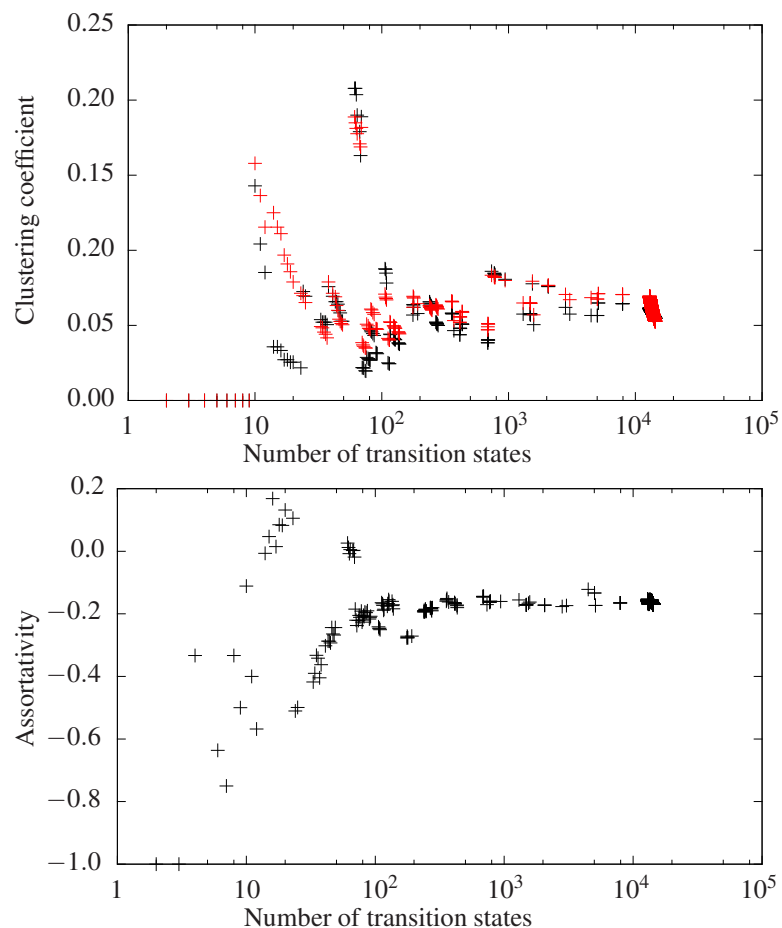




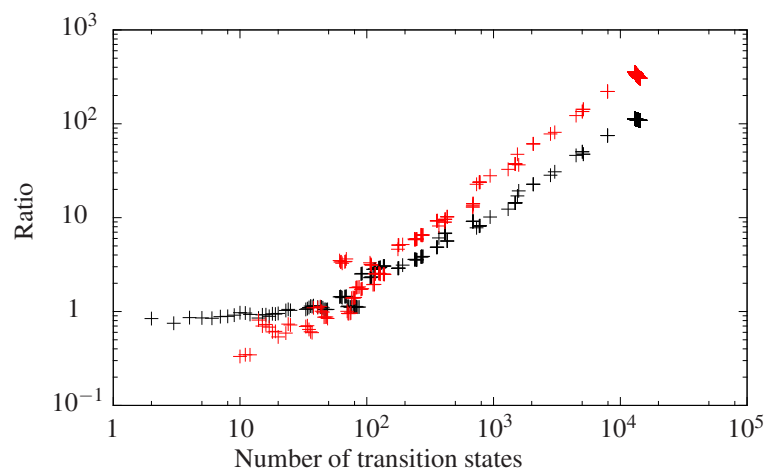
**Fig. 8** Top: the variation of the average clustering coefficient (black) and transitivity (red) as the largest connected component of the OTP database grows. Bottom: the corresponding plot for the assortativity.



**Fig. 9** The ratio of average shortest path length for the largest connected component of the OTP database to that of the equivalent random graph (black) and the corresponding ratio for the transitivity (red).



**Fig. 10** Top: variation of the average clustering coefficient (black) and transitivity (red) as the largest connected component of the BLJ database grows. Bottom: the corresponding plot for the assortativity.



**Fig. 11** The ratio of average shortest path length for the largest connected component of the BLJ database to that of the equivalent random graph (black) and the corresponding ratio for the transitivity (red).